

Age Harmonization in COVerAGE-DB

Tim Riffe

10/28/2020

Efficient Estimation of Smooth Distributions From Coarsely Grouped Data

Silvia Rizzi , Jutta Gampe, Paul H. C. Eilers [Author Notes](#)

American Journal of Epidemiology, Volume 182, Issue 2, 15 July 2015, Pages 138–147,
<https://doi.org/10.1093/aje/kwv020>


Published: 16 June 2015 **Article history** ▼

A correction has been published:

American Journal of Epidemiology, Volume 185, Issue 4, 15 February 2017, Page 316,
<https://doi.org/10.1093/aje/kww225>



PDF

 Split View

 Cite

 Permissions

 Share ▼

Abstract

Ungrouping binned data can be desirable for many reasons: Bins can be too coarse to allow for accurate analysis; comparisons can be hindered when different grouping approaches are used in different histograms; and the last interval is often wide and open-ended and, thus, covers a lot of information in

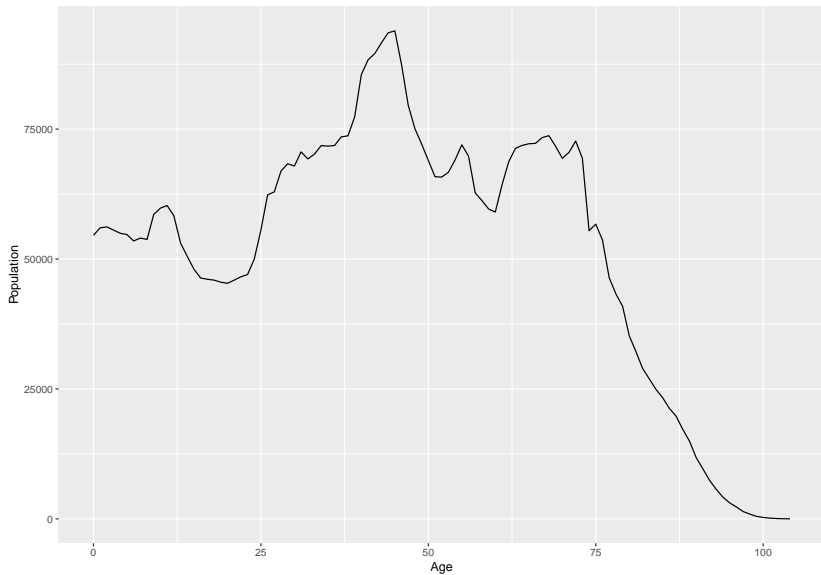
Features

- ▶ Implemented in `ungroup` R package
- ▶ If offsets (population) used, then smoothness penalty happens at the rate scale
- ▶ Smoothness (of implied rates) controlled by a single parameter, λ
- ▶ COVerAGE-DB uses $\lambda = 100000$, very smooth, good for old ages
- ▶ Can also auto-select λ
- ▶ Tim's fav graduation method (comparing w Sprague, Beers, splines)

Performance examples

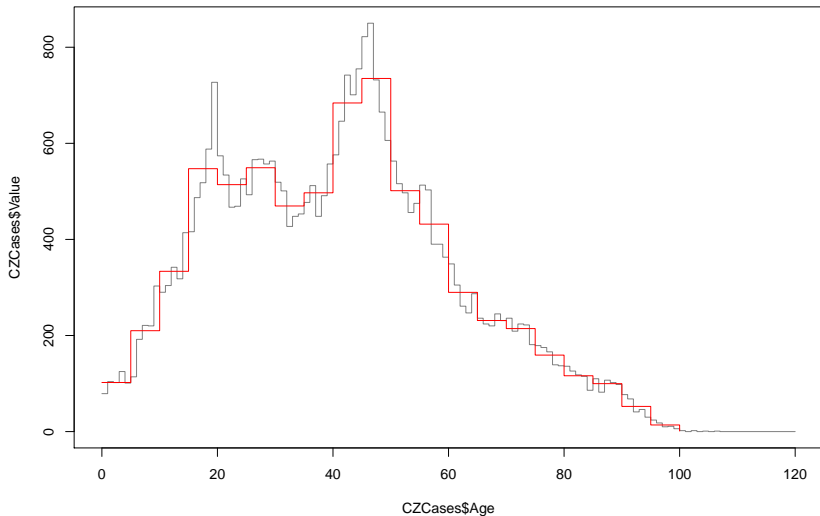
- ▶ Show performance based on full-information case
- ▶ CZ has single-age offsets and confirmed cases (Oct 10, 2020) (some noise)
- ▶ Group cases to 5, 10, 20 year age groups
- ▶ Use PCLM to get back to single ages, compare
- ▶

Population structure, Czechia females



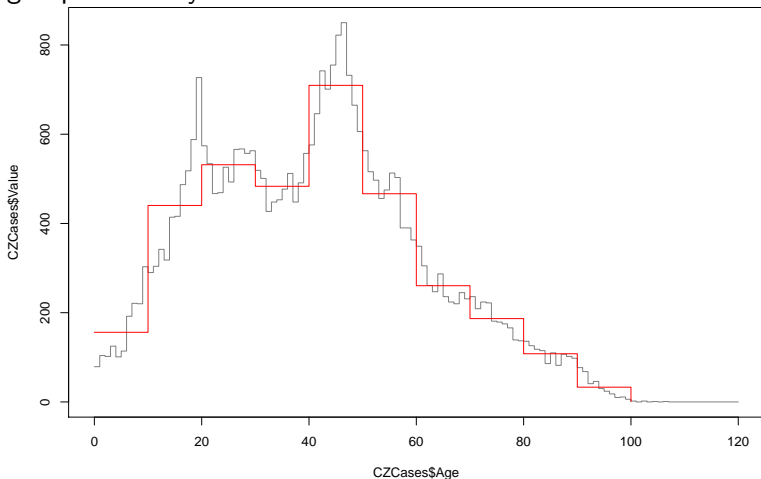
Cases (1 & 5 year intervals)

- ▶ gray = orig
- ▶ grouped to 5-yr intervals



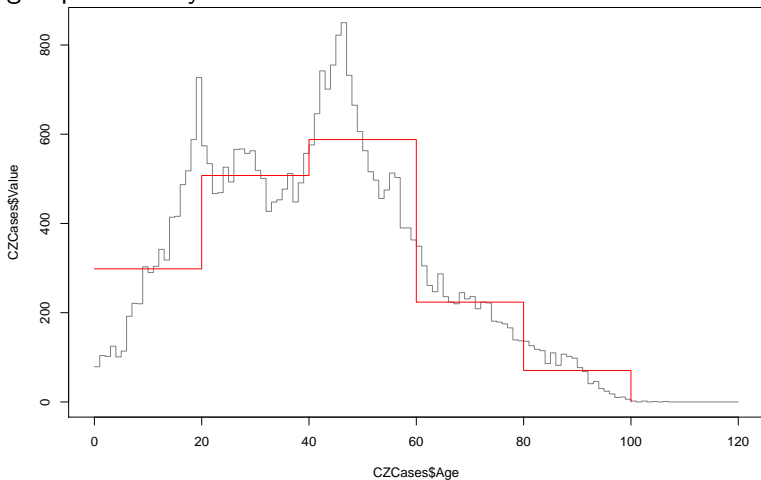
Cases (1 & 10 year intervals)

- ▶ gray = orig
- ▶ grouped to 10-yr intervals



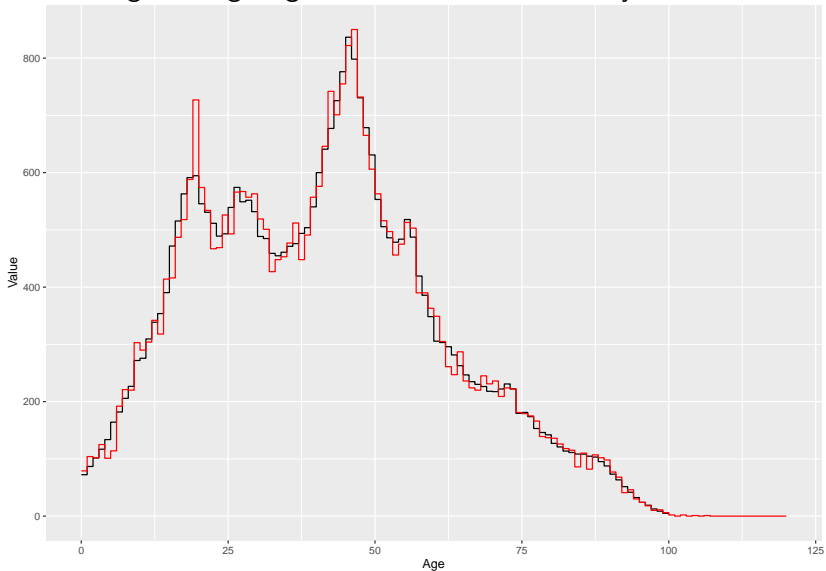
Cases (1 & 20 year intervals)

- ▶ gray = orig
- ▶ grouped to 20-yr intervals



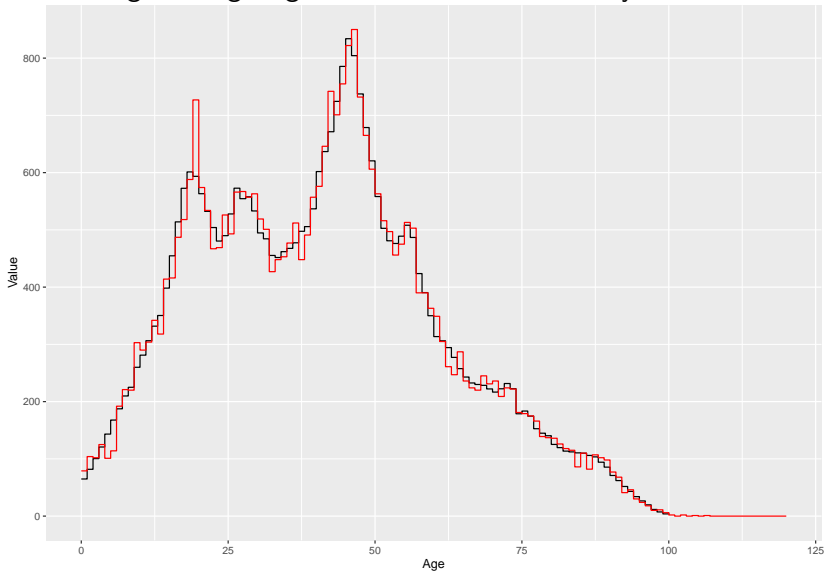
Performance comparisons 5 \rightarrow 1, very smooth

Red = original single ages Black = PCLM from 5-year intervals



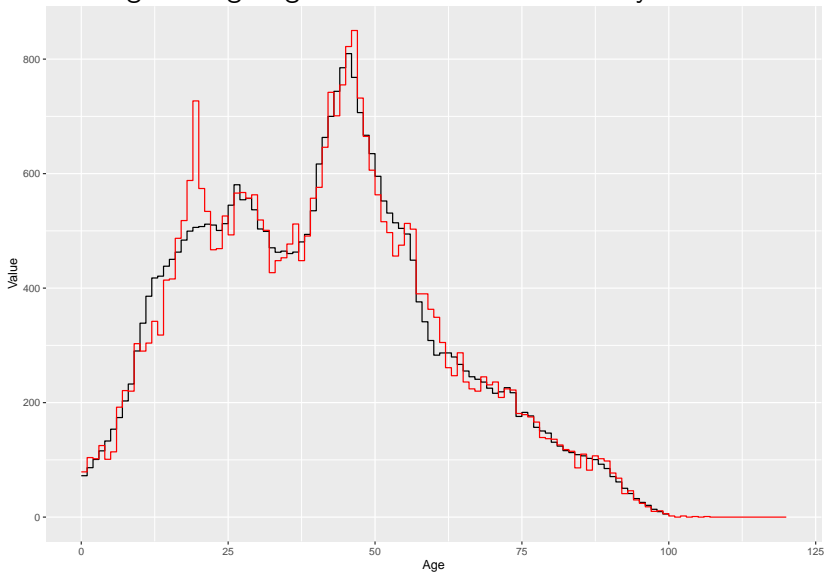
Performance comparisons, 5 \rightarrow 1, less smooth

Red = original single ages Black = PCLM from 5-year intervals



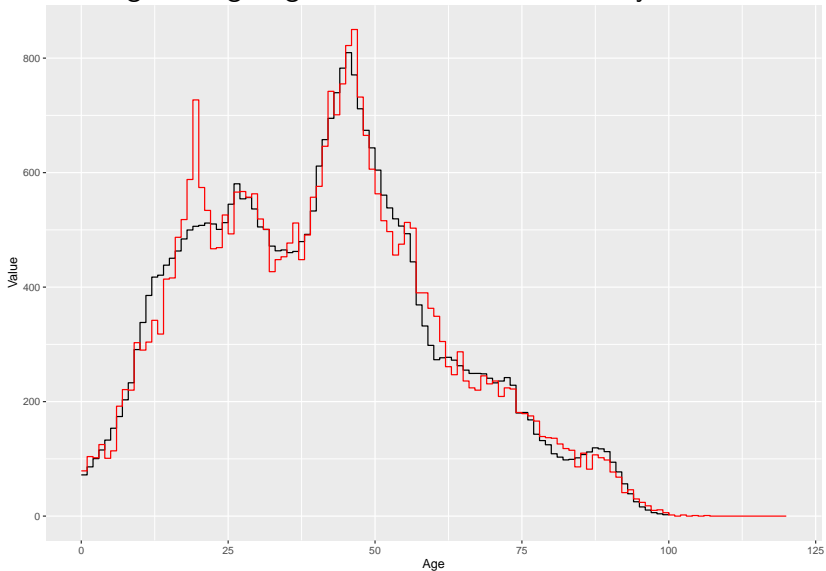
Performance comparisons 10 -> 1, very smooth

Red = original single ages Black = PCLM from 10-year intervals



Performance comparisons 10 \rightarrow 1, less smooth

Red = original single ages Black = PCLM from 10-year intervals



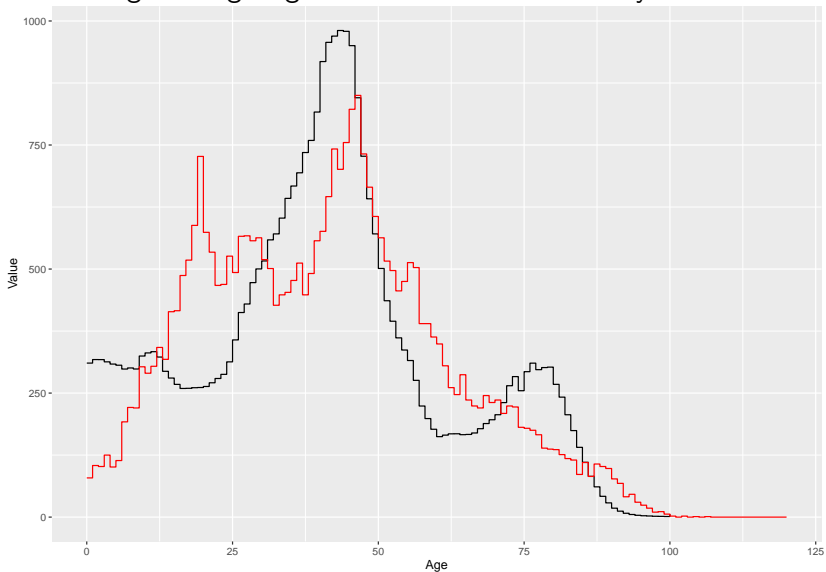
Performance comparisons 20 -> 1, very smooth

Red = original single ages Black = PCLM from 20-year intervals



Performance comparisons 20 \rightarrow 1, less smooth

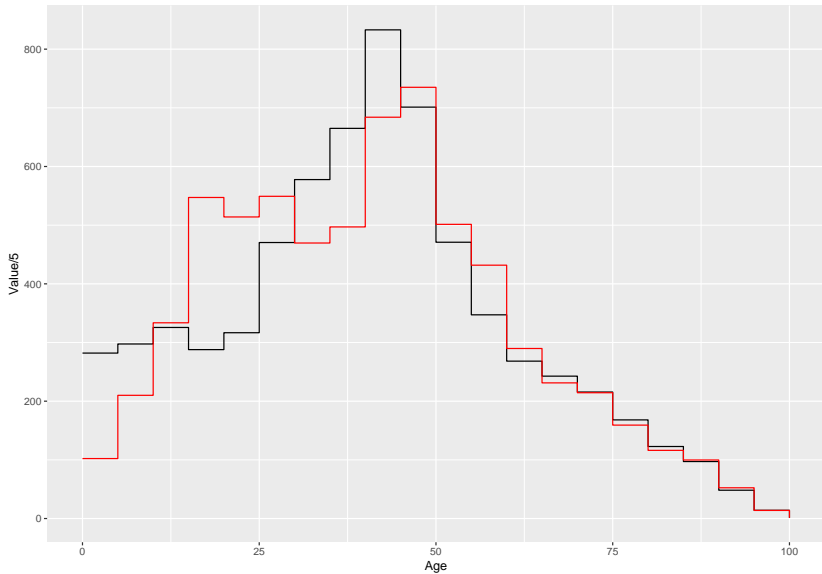
Red = original single ages Black = PCLM from 20-year intervals



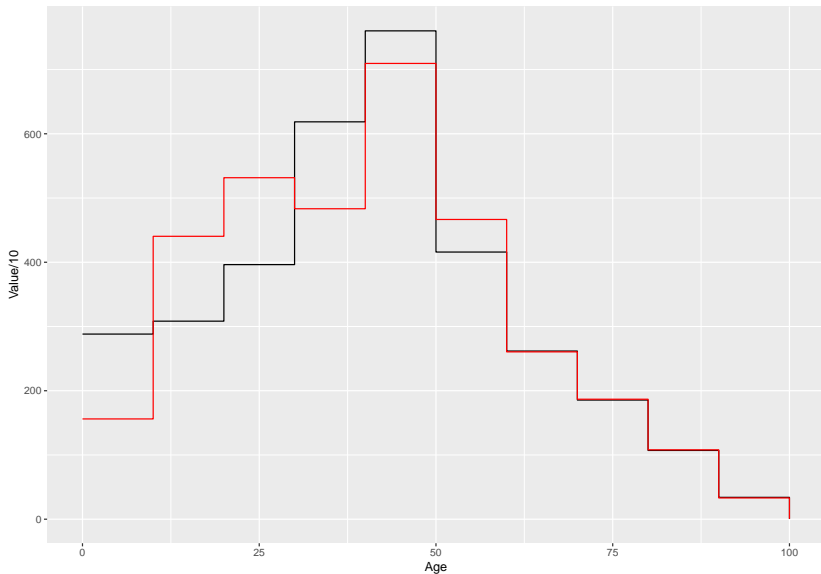
Wide age groups look pretty bad, right?

- ▶ Less bad when output is delivered in 5-year age groups
- ▶ Less bad when offsets available
- ▶ Less bad when related to population, deaths, or similar

Example: $20 \rightarrow 1 \rightarrow 5$, compare w orig 5



Example: $20 \rightarrow 1 \rightarrow 10$, compare w orig 10



Comments

- ▶ Age patterns of cases appear harder to match than deaths
- ▶ Results slightly better with auto-select λ (not shown), but runs slow
- ▶ COVerAGE-DB might change protocol, this is just current fav
- ▶ Thanks!
- ▶ R Markdown for this pres https://github.com/timriffe/covid_age/blob/master/Talks/AgeHarmonization.Rmd
- ▶ Questions to Tim: riffe@demogr.mpg.de